DOCUMENT RESUME

ED 340 037 CS 213 116

AUTHOR Brand, Alice G.
TITLE Constructing Tasks for Direct Writing Assessment: A Frontier Revisited.
PUB DATE 91
NOTE 21p.
PUB TYPE Viewpoints (Opinion/Position Papers, Essays, etc.) (120) -- Guides - Non-Classroom Use (055)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Higher Education; Secondary Education; *Test Construction; Test Validity; *Writing Evaluation; *Writing Tests
IDENTIFIERS *Writing Tasks; *Writing Test Prompts

ABSTRACT

Since the 1980s, composition studies have considered the steps to be taken before assessment: designing the test essay question. While large-scale assessment has little control over writing variables (such as students' learning styles, their reading ability, and their interpretation of the topic), the content or the topic of the writing prompt and the procedures for writing can be manipulated, or at least acknowledged. In 1984, A. C. Purves and his colleagues classified 15 dimensions of the writing task that could be brought to bear on a writing prompt. This list provides a sense of the options available even though only a few can be dealt with on any one test. Ways to overcome the problem of choosing an appropriate topic include: select a topic of more or less equal familiarity to all students; generate a pool of topics of similar difficulty; supply the information for the topic; and allow students to write about whatever subjects they care about and/or believe they have the most knowledge of. A study comparing three topical structures in the writing prompt found that their utility depended upon purpose. In the interests of consistency, topics across an assessment and the accompanying instructions should share several of the characteristics that reduce the risk of interpretations too different from the intentions of the task. Four questions can guide broad decisions about writing assessment design and furnish test developers with the raw materials to create fair and meaningful writing prompts: (1) To what degree will the assessment be content specific? (2) What are the rhetorical demands of the task? (3) What topical structure and level of specification should be used? and (4) How should the prompt look on the page? (A table listing the dimensions of a writing assignment is included; 23 references are named.) (RS)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Alice G. Brand
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Alice G. Brand
SUNY Brockport

ED340037

Constructing Tasks for Direct Writing Assessment
A Frontier Revisited

Large-scale outcomes assessment has become a commonplace enterprise across the country. And writing is no exception. Just when assessment designs and scoring methods have become virtually axiomatic, a relatively young avenue of inquiry is uncovered: that into the writing task itself. For almost a decade now composition studies has been quietly but seriously looking at the step before assessment that brings it all about: designing the test essay question. And this is subject to more question. Let me first summarize what the field generally agrees on.

1. The emergence of writing as a discipline gave impetus to a movement in college writing programs away from multiple choice and short answer tests to direct measures-- those that call for writing samples.
2. Testing with writing samples is equally efficient and cost effective as objective, indirect measures.
3. Furthermore, a fairly reliable picture of students' competencies can be drawn from writing completed in anywhere from 20 minutes (like the SAT) to two hours (White 1985).
4. Direct measures of writing can serve several functions: admission, placement, equivalency, exit/barrier, and program assessment. They can thus certify student proficiency in writing as freshmen, as rising juniors or transfers, as

CS213116

BEST COPY AVAILABLE

writers in their major, and/or as graduating seniors.

5. The two major designs of scoring single essays are Analytic and General impression scoring, the latter of which may be further subdivided into Holistic and Primary Trait.

Analytical Scoring is a trait-by-trait analysis of features important to any piece of discourse. The features: ideas, organization, wording, flavor, usage, punctuation, spelling, and handwriting are scored on a 5-point scale, with the first four subskills receiving double weighting because of their importance to the success of the essay (Diederich 1974).

General Impression scoring consists of two major types: Holistic scoring, which originated through the Educational Testing Service, and Primary Trait scoring, which was developed by Richard Lloyd-Jones (1977) for the National Assessment of Educational Progress. Holistic Scoring is based on readers' overall impressions of the effectiveness of a piece of writing. The scoring guide consists of a numerical rating spread generally over 4 or 6 points which is accompanied by a description of the features that characterize each rating. Primary Trait Scoring focuses on specific rhetorical features or traits of a given piece of writing. As a form of holistic scoring, this approach identifies the mode of discourse to be evaluated (for example, description, comparison, analysis of a process, argument) and trains readers to render holistic judgments

according to the unique features of that form. For example, an argument would be judged on the quality of its evidence, a description on its sensory detail. The most important advantage of primary trait scoring over holistic scoring is that the primary trait approach defines with greater precision and exclusiveness the criteria used in holistic scoring.

6. Although the pre-post essay has been popular direct measure of writing skills, the shift from a written product to a writing process pedagogy has meant a growing interest in portfolio assessment. Briefly, in portfolio assessment a collection of work written over a given period is evaluated. Depending on how portfolio assessment is implemented, it may also include intermediary drafts of that work.

Major Task Variables

However the collecting and evaluating of writing samples is undertaken, task construction is particularly confusing and thus warrants particular attention. Brossell (1986) divides the array of task variables into three groups: writing, procedural, and topic.

I will leave aside the plethora of writing (more accurately called writer) variables over which large-scale assessment has little control: students' social and cognitive makeup and their affective predispositions. And within those vast domains: students' learning styles, their reading ability, their interpretation of the topic, their expertise on it, their

proficiency with the language, their affective responses to topics and testing conditions, and of course, their intrinsic motivation--a crucial intervening variable well outside of institutional jurisdiction.

But the content or topic of the prompt and the procedures for writing about it are another story. Like writing variables, procedural and content variables appear legion. However, unlike writing variables, procedural and content variables can be manipulated. Or if they cannot be manipulated at least they can be acknowledged. At the risk of stating the obvious, these dimensions can make a dramatic difference in fluency, quality, and total errors (Smith et al. 1985). What specialists are just beginning to learn about is how.

Dimensions of Task Construction

In 1984 Purves and his colleagues classified 15 dimensions of the writing task that could be brought to bear on a writing prompt (see Table 1 below), a breakdown if you will of Brossell's content and procedural variables. In terms of empirical research

---

Insert Table 1 About Here

---

the possibilities outlined in Purves's taxonomy overwhelm. In terms of test essay construction the list provides a sense of the options available even though perhaps only a few can be reasonably dealt with during any one test cycle. To some extent our intuitions can guide decisions as to which items are

Dimensions of a Writing Assignment

A. Instruction
   1. Topic only
   2. Topic and content cues

B. Stimulus
   1. Unspecified
   2. Word, phrase, or sentence
   3. Test (non-literary)
   4. Test (literary)
   5. Photograph
   6. Music
   7. Other media
   8. Realia (field trip)
   9. Mixed

C. Cognitive Demand
   0. Unspecified
   1. Reproduce facts
   2. Reproduce ideas
   3. Organize, reorganize events
   4. Organize, reorganize visual images, ideas, etc.
   5. Invent, generate, evaluate

D. Purpose
   0. Unspecified
   1. To learn
   2. To convey, signal
   3. To inform
   4. To convince, persuade
   5. To entertain
   6. Multiple

E. Role
   0. Unspecified
   1. Write as self (journal, monologue, etc.)
   2. Write as self (autobiography)
   3. Write as self (detached observer)
   4. Write as other (assumed persona)
   5. Write as other (detached author)

F. Audience
   0. Unspecified or teacher
   1. Self
   2. Specified individual (addressee of letter)
   3. Specified group
   4. Classmates
   5. General Public
   6. Other
   7. With pre-writing activities (>120 minutes)
   8. Impromptu homework
   9. With pre-writing homework

G. Content
   0. Unclassified, multiple
   1. Self
   2. Family, relatives, home
   3. Peers
   4. Specific individual (non-literary)
   5. School, schooling, education
   6. Leisure activities (sports, hobby, reading)
   7. Community
   8. School subject: mother tongue, literature
   9. School subject: other
   10. Social, political, cultural, economic issues: general
   11. Science, man and nature, ecology: general
   12. Psychology, philosophy, religion, ethics: general
   13. Other

H. Rhetorical specification
   0. Unspecified or multiple
   1. Note, informal letter
   2. Formal letter (e.g. business letter)
   3. Resume, summary, paraphrase
   4. Narrative
   5. Description
   6. Exposition, analysis, definition, classification
   7. Narrative, descriptive, exposition, with evaluative comment
   8. Argument, with evaluation and comment
   9. Literary genre
   10. Advertisement, media
   11. Journal writing

I. Tone, style
   0. Not specified
   1. Overall tone, style specified
   2. Particular stylistic devices specified
   3. Overall tone and style and stylistic devices specified

J. Advance preparation
   0. None, unspecified
   1. Class discussion
   2. Group discussion
   3. Collecting materials
   4. Previous work studied

K. Length
   0. Unspecified
   1. One paragraph (<150 words)
   2. 2-5 paragraphs (<2 pages)
   3. 2-5 sides (300-700 words)
   4. >5 sides

L. Format
   0. Unspecified
   1. Specified as to use of space on page
   2. Specified as to conventions
   3. Other specifications

M. Time
   0. Unspecified
   1. Impromptu (<30 minutes)
   2. Impromptu (30-59 minutes)
   3. Impromptu (60-120 minutes)
   4. Impromptu (>120 minutes)
   5. With pre-writing activities (<60 minutes)
   6. With pre-writing activities (60-120 minutes)

N. Draft
   0. Unspecified
   1. One draft only
   2. One draft with corrections
   3. Rough draft and revision with second copy
   4. Rough draft and revision to final edited copy

O. Criteria
   0. Unspecified
   1. Grammar and orthography primarily
   2. Style and appropriateness primarily
   3. Organization and presentation primarily
   4. Ideas and content primarily
   5. Neatness primarily
   6. Combinations of above
   7. Other criteria

Purves et al. 1984.

BEST COPY AVAILABLE

sufficiently relevant to warrant their appearance in the prompt, which are worth controlling for, and which may be gainfully ignored. Indeed, some dimensions may be for all practical purposes out of our hands: the 50-minute class session, the number of drafts, the amount of advanced preparation-- and where there is, the method of authentication. Other components seem correctly inconsequential. For example, small changes in phrasing seem to produce no significant change in essay scores (Brossell and Ash 1984; Brossell and Hoetker 1989; Greenberg 1981). Still other components may even surprise. For example, time does not appear to be an important variable. But only a certain kind of time (and under certain testing conditions). One or two hours does not seem to make a significant difference in essay scores (Dunham and Coppersmith 1987; Polin 1980); furthermore, as we know from ETS, even a 20-minute essay can be uniquely telling.

Then some dubious options, to wit: audience. Despite our abiding fascination with audience, specifying an audience for a piece of writing seems to make a bigger difference to test developers than to average test takers (Huot 1990). Prompts with high audience specifications yield high scores under some circumstances (Brossell qtd. in Greenberg 1986, 174); under other circumstances they don't seem to matter (Redd-Boyd and Slater 1989; Ruth and Murphy 1988, 140). Also inconclusive is the relationship between writing quality and the rhetorical modes such as narration, definition, comparison/contrast, and analysis--called rhetorical specification in Purves's taxonomy. While it

has been conventional wisdom that students writing personal experiential papers generally perform better than those writing more impersonal, academic essays (Hoetker and Brossell 1986) and that an argument is more difficult to respond to than a narrative (Applebee and Brossell 1985), other research has found either the opposite (Hoetker 1982) or no statistical differences in quality between personal experiential writing and highly cognitive (more academic) writing (Greenberg 1981; Hoetker and Brossell 1989; Quellmalz et al. 1982). This last issue is not particularly newsworthy in and of itself, except for the fact that there is no reliable system for distinguishing between rhetorical modes of development anyway. What's more, in both real world and academic writing, pure modes rarely exist. Narratives typically include description; of necessity, classifications must compare, contrast, and define; arguments must in some way describe and analyze. Good writing does these things.

But nowhere in test question design is care more critical than in the overall presentation of the writing prompt. By this I mean

1. the task content or subject matter; and
2. the task specification or structure of the prompt.

Task Content

The selection of topic is heady stuff, ranging from the self as noted in Purves's system to the world of ideas. Of course, different topics demand different knowledge bases and vocabularies, which introduce a variety of syntax and essay

length. But more important is the fact that no matter what subject is chosen, some students will be better equipped to write about it than others (Hoetker and Brossell 1986, 329). One way to overcome the problem of unequal familiarity with a topic is to select one of more or less equal familiarity: For example, however different the experience, by and large college students have in common the high school experience, some social experience, college orientation, course registration, a science lab.

Another way to solve this problem is to generate a pool of topics of similar difficulty -- considerably harder to do. Is writing about noise pollution the rhetorical equivalent of writing about air pollution? What activities correspond intellectually to writing about baby sitting for a grandparent? Still another way to enhance reliability of test topics is to supply the information, a familiar option for test developers and instructors. In the Smith et al. study summarized below topical structures providing written passages were examined. Along with an open-ended essay, SUNY Brockport's composition exit exam attempts to do this with a two-page reading selection provided at the exam. Some schools send their students home with considerable reading matter several weeks before an exam. Others have begun basing their upper level competency tests on sustained reading on a single topic before actually writing the test. The knowledge base is controlled to some extent but at a price: Reading passages complicates testing logistically because of its claim on

time. More important, because reading means interaction with texts, which means comprehending and interpreting them, it introduces complicated sociocognitive and affective variables.

Still a fourth solution to the prcblem of varying familiarity of students with test topics is to allow students to write about whatever subjects they care about and/or believe they have most knowledge of. These topics may be considered "content-fair," that is, neutral or open. Such a model guided a study designed to test this method undertaken by Hoetker and Brossell in 1986. A "content-fair" essay prompt was generated for rising juniors in the State of Florida's College-Level Academic Skills Test. A "master topic" was based on "frame" sentences with individual topics produced by substituting different words or phrases in specified blanks in the master pattern. The structure of a formal definition provided the master pattern:

1. A category name (a noun or noun phrase) - e.g., "A flute";
2. A linking verb - "is";
3. A class - "a keyed woodwind instrument";
4. Differentiating attributes - "which has a side hole over which wind is blown, etc."

The test essay topic consisted of the class specification (3) and two differentiating attributes (4). Students could then choose a specific example of the category or noun (1) that completed the definition. For example, a topic might read:

"A novel/ which many students read/ that may affect them significantly."

Students are then free to identify a novel which they know about and like--or dislike. Moreover, they can supply their own definition of "significant effects." This so-called "topic-generating machine" can then produce any number of syntactically and semantically similar topics. Each topic allows students to supply their own specific subject. The sample topic can also be modified:

A book/that many students read/ that may affect them harmfully.
A book/ that many students read/ that may affect them beneficially.
A course/ that many students take/ that may affect them in important ways.
A habit or belief/ that many students acquire in college/ that affects their life [sic] in important ways.

If another category or other differentiating attributes are substituted, another set of related subjects can be generated. For example:

A common practice/ in American schools/ that should be changed.
A common practice/ in American schools/ that inhibits learning.
A common practice/ in American colleges/ that should be changed.
A common practice/ among instructors/ that does not achieve what it is intended to achieve.
A common practice/ in American elections/ that does not serve the interests of the public.

(Hoetker and Brossell 1986, 329-331)

To the extent that strong and highly verbal students write better about abstract subjects while average and basic students write more comfortably about people, places, and things, Hoetker and Brossell recommend that two topics be developed--one with an

abstract noun and one with a concrete noun--between which students may choose. Clearly, students can make bad choices. They may find themselves rudderless, unimaginatively describing the school cafeteria food--an overworked subject at best. Or they may find themselves entirely out of their depths trying to analyze the world monetary fund in three pages, as one of my students recently tried to do.

Topic Specification and Structure

Just how much and what kind of information should be provided in the writing prompt should also come under scrutiny. How much language should be devoted to content? How much to instruction? What can be taken for granted? What must be specified? There are no easy answers here. Capable of producing equally fatal or stunning responses are prompts of one and a half single-spaced page of virtually pure instructions, a fictional scenario with a directive to "discuss it, " or a simple quotation with minimal instructions like "Americans would rather watch a sport than participate in it. Do you agree or disagree? Take a stand."

Three topical structures were considered in one investigation (Smith et al. 1985): The first was an open structure in which students were given a broad topic, asked to draw on personal knowledge and experience and to generalize from them. The next was a response structure with one reading passage which students were asked to read and respond to in some way. The last was a response structure with more than one reading

passage in which students were asked to describe what each author wrote about and generalize from them both.

No structure was found without merit. Rather, Smith and his colleagues found utility depended on purpose. The open structure was useful when high ability writers were to be distinguished from average and low ability writers because superior writers are more able to marshall and order their evidence quickly (83). If lower ability writers were to be identified (as in placement tests), the multiple reading response structure was recommended because it elicits the greatest length and the greatest number of errors (83-84). For exit exams (e.g., essays used to determine who is proficient, who may be waived from a course, or who may move to a more advanced one) the recommended topic structure seems to depend on how cautious one wishes to be (84). Open topic structures, for example, might prevent marginal students from moving on to new courses, whereas the multiple passage response structure apparently allows some less able writers to pass (84). Both the open topical structure and the multiple reading structure generally yielded greatest length. At the same time students seemed to make most errors with the multiple reading structure because the prompt made the highest cognitive demands on them--the amount of evidence that students need to process (82).

Reading only one passage seemed least appropriate overall because it did not clearly distinguish students at various proficiency levels. Fewer errors were committed with the one

passage structure because students borrow text, in addition to which this structure tended to suppress fluency. However, the one-passage option was not to be ignored, maintained Smith. In some situations it is desirable because it produces the shortest essays that can be read and rated the fastest. In addition, the one-passage prompt provides less chance that raters will be influenced by length, yet it will sort students (84).

Until the relationship between topic specification and writing quality is fully elaborated, some of the precautions Hoetker and Brossell considered in 1986 should probably be taken. In the interests of consistency, topics across an assessment and the accompanying instructions should share several characteristics that reduce the risk of "interpretations too different from [the] intentions" [of the task] (329). A controlled pre-post assessment design should particularly "be cast in a common syntactic pattern and ... should share semantic elements as well" (329). The more words in a test essay prompt, the more semantic relationships students must negotiate, and the greater the risk of misunderstanding. All things being equal, shorter or "frame" questions are preferred to fully specified ones because there is "no surer way to reduce misreadings [of an assignment] than to reduce the number of words that had to be read and interpreted" (329). Brief prompts are also preferred because students are better able to produce their own ideas (Hoetker 1982). However, for students suffering from idea-anemia, prompts that are moderately specified rather than open-ended

elicit more focused and better organized essays as well as high mean scores (Brossell 1983). This is particularly important when time is at a premium (Applebee and Brossell 1985).

A word about audience specification. By-words audience and purpose have come almost full circle from implicit specification to hyper-specification and back to moderate specification. The field seems to be coming to its senses here. Realistic social or professional purposes create useful rhetorical contexts for writing, but highly bogus scenarios or cases have fallen from grace. No more pretending readers are visitors from a small planet. No more time warp to pre-historic Budapest. Gone are the days of fabricating an audience of Rosicrucians. Besides which students know who their real audience is: by and large, their instructors--intelligent and largely informed professionals. Instructions need to be clear and honestly descriptive of the academic reality. This is way we introduced our placement essay:

> This pretest gives your instructor a sense of your ability to write academic prose under impromptu conditions. Because this pretest resembles the kind of writing done in college, it should help you focus on the writing skills you may need to work on this semester.

The prompt which followed served us well, giving students some choices as well as enabling us to distinquish broadly among several levels of writing skill:

Choose one of the topics below and write an essay of several paragraphs that either explains your subject or takes a position about your subject and supports it.

A. A book that my classmates and I have read that has influenced us.
B. A common practice in high school that should be changed.
C. A national or international situation that may affect my family and/or my friends.

Your essay will be evaluated according to 3 criteria: your ability to

A. limit and focus your topic (underline your controlling idea or thesis);
B. develop concrete and specific detail to support your facts and opinions; and
C. use standard usage, grammar, and punctuation.

You will have this class session to plan, write, and proofread your essay. Feel free to use the first page or two to plan. Please write on only one side of each page and skip a space between each line.

For a job that seems deceptively simple, devising a writing task should also not get obsessively picayune. In our own work at SUNY Brockport we were far more worried about all 15 dimensions than we needed to be. The one recurring theme in the literature and confirmed in my own experience is that even when assignment variables are tested carefully, students' written performance is far more impervious to change than has been expected. In other words, it seems to take substantial change in the test question to produce palpable change in written response.

But this does not mean we should abrogate our responsibility for reasoning through the assignment construction process. With what little we know some priorities can still be established.

The questions below not only guide broad decisions about writing assessment design but also furnish test developers with the raw materials to create fair and meaningful writing prompts.

1. To what degree will the assessment be content specific?

   What, if any, subject(s) will be used? How appropriate is it (are they)?

   What prior knowledge is required?

   Is it relevant to students' lives, their majors, their course work?

   Will it evoke stereotypical thinking? Will it threaten personal, ethnic, or cultural values? Will it be relatively free of bias?

2. What are the rhetorical demands of the task?

   What is the intention of the task? How should it be stated?

   Which, if any, rhetorical modes of development will be tested? [1]

   How does the choice of subject affect the choice of rhetorical mode?

   Should the prompt specify an audience? How should it be stated?

3. What topical structure and level of specification should be used?

   What percentage of language should be given over to content and what to instruction?

   How should the prompt be worded? Verbs? Prose labels (essay, argument, persuasive essay)?

   What should be the length and complexity of the sentence(s)?

   To what extent can the words and structures used in the answer be borrowed directly from the prompt or text?

   What are the benefits and liabilities of the chosen

---

[1]White suggests at least 2 (qtd. in Greenberg et al. 1986, 67).

structure and specification level?

4. How should the prompt look on the page? Is there sufficient white space? Is it 100% readable?

Most heartening through all this is that some of these matters are gradually stabilizing. Instructors and test makers are already beginning to treat writing task construtions with greater sanity. No longer are they instantly and sometimes mindlessly responsive to every piece of research published on prompt construction. The literature provides a number of justifiable choices. Test designers are recognizing many rights as well as at least several wrongs with each. They are reasoning their way through to a prompt given their particular academic context. As the field moves on to other, more interesting issues, such choices need to be given more time to demonstrate their powers of discrimination.

Works Cited and Consulted

Applebee, A. and G. Brossell. 1985. "Research on Writing Assessment." Notes from the National Testing Network in Writing V: 10.

Brossell, G. 1983. "Rhetorical Specification in Essay Topics." College English 45: 165-173.

---. "Essay Test Topic Development." 1986. Notes from the National Testing Network in Writing VI: 6-7.

Brossell, G. and Ash, B.H. 1984. "An Experiment with the Wording of Essay Topics." College Composition and Communication 35: 423-425.

Diederich, P. 1974. Measuring Growth in English. Urbana, IL: National Council of Teachers of English.

Dunham, D. and S. Coppersmith. 1987. "Content-Fulfillment and Creativity: A Problem for Testing." Fifth Annual Conference on Writing Assessment. Atlantic City, NJ, 7 April.

Greenberg, K. 1981. The Effects of Variations in Essay Questions on the Writing Performance of CUNY Freshmen. New York: The City University of New York Instructional Resource Center.

Greenberg, K. L., H. S. Wiener, and R. Donovan. 1986. Writing Assessment: Issues and Strategies. White Plains, NY: Longman.

Hoetker, J. "Essay Examination Topics and Students' Writing."

1982. College Composition and Communication 33: 377-392.

Hoetker, J. and G. Brossell. 1986. "A Procedure for Writing Content-Fair Essay Examination Topics for Large Scale Writing Assessments." College Composition and Communication 37: 328-335.

---. 1989. "The Effects of Systematic Variations in Essay Topics on the Writing Performance of College Freshmen." College Composition and Communication 40: 414-421.

Huot, B. 1990. "The Literature of Direct Writing Assessment: Major Concerns and Prevailing Trends." Review of Educational Research 40: 237-263.

Lloyd-Jones, R. 1977. "Primary Trait Scoring." Evaluating Writing: Describing, Measuring, Judging. eds. Charles Cooper and Lee Odell. 33-66. Urbana, IL: NCTE.

Polin, L.G. 1980. Effects of Time and Strategy Use on Writing Performance. Los Angeles: UCLA Center for the Study of Evaluation.

Pollitt, A., C. Hutchinson, N. Entwistle, and C. De Luca. 1985. What Makes Exam Questions Difficult? Edinburgh: Scottish Academic Press.

Purves, A. C., A. Soter, S. Takala, and A. Vahapassi. 1984. "Towards a Domain-Referenced System for Classifying Composition Assignments." Research in the Teaching of English 18: 385-416.

Quellmalz, E. S., F. Capell, and C.P. Chou. 1982. "Effects of Discourse and Response Mode on the Measurement of Writing

Competence." Journal of Educational Measurement 19: 241-258.

Redd-Boyd, T. M. and W. H. Slater. 1989. "The Effects of Audience Specfication on Undergraduates' Attitudes, Strategies, and Writing. Research in the Teaching of English 23: 77-108.

Ruth, L. and S. Murphy. 1988. Designing Writing Tasks for the Assessment of Writing. Norwood, NJ: Ablex.

Scholastic Aptitude Tests. n.d. Princeton, NJ: Educational Testing Service.

Smith, W. L., G. A. Hull, R. E. Land, Jr., M. T. Moore, C. Ball, D. E. Dunham, L. S. Hickey, and C. W. Ruzich. 1985. "Some Effects of Varying the Structure of a Topic on College Students' Writing." Written Communication 2: 73-89.

White, E. M. 1990. "Language and Reality in Writing Assessment." College Composition and Communication 41: 187-200.

---. Teaching and Assessing Writing. 1985. San Francisco: Jossey-Bass.